A Novel Equivalent Model of Active Distribution Networks Based on LSTM

Zheng, Chao; Wang, Shaorong; Liu, Yilu; Liu, Chengxi; Xie, W.; Fang, Chen; Liu, Shu

Published in:
IEEE Transactions on Neural Networks and Learning Systems

DOI (link to publication from Publisher):
10.1109/TNNLS.2018.2885219

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

A Novel Equivalent Model of Active Distribution Networks Based on LSTM

Chao Zheng, Shaorong Wang, Yilu Liu, *Fellow, IEEE*, Chengxi Liu, Wei Xie, Chen Fang, and Shu Liu

***Abstract*— Dynamic behaviors of distribution networks are of great importance for the power system analysis. Nowadays, due to the integration of the renewable energy generation, energy storage, plug-in electric vehicles, and distribution networks turn from passive systems to active ones. Hence, the dynamic behaviors of active distribution networks (ADNs) are much more complex than the traditional ones. The research interests how to establish an accurate model of ADNs in modern power systems are drawing a great deal of attention. In this paper, motivated by the similarities between power system differential algebraic equations and the forward calculation flows of recurrent neural networks (RNNs), a long short-term memory (LSTM) RNN-based equivalent model is proposed to accurately represent the ADNs. First, the adoption reasons of the proposed LSTM RNN-based equivalent model are explained, and its advantages are analyzed from the mathematical point of view. Then, the accuracy and generalization performance of the proposed model is evaluated using the IEEE 39-Bus New England system integrated with ADNs in the study cases. It reveals that the proposed LSTM RNN-based equivalent model has a generalization capability to capture the dynamic behaviors of ADNs with high accuracy.**

***Index Terms*— Deep learning, dynamic behaviors, load modeling, long short-term memory (LSTM), measurement-based approach, recurrent neural network (RNN).**

I. Introduction

DUE to the factor that load behaviors have noticeable effects on the dynamic response of power systems, load modeling is one of the most fundamental aspects for power system planning, analysis, operation, and control [1]–[5]. Kinds of the literature show that if the load model failed to accurately represent the dynamic characteristics of distribution systems, incorrect conclusions would be drawn in the stability analysis of power systems, such as voltage stability assessment and dynamic security analysis [6]–[10]. For example, the simulation results did not agree with the disturbance recordings from the Western Electricity Coordinating Council (WECC) blackout in 1996, when the models from the standard WECC database were adopted into the simulations [11]. Hence, researchers and engineers are fully motivated to develop an accurate load model for power system studies.

Nowadays, the distribution networks gradually evolve from passive systems to active ones, because of the integration of renewable energy generation, energy storage, and plug-in electric vehicles in distribution networks. Moreover, due to the aforementioned factors and the wide applications of power electronic converters, the dynamic behaviors of the distribution networks are more complex than the traditional ones. Conventional load models are difficult, even impossible, to represent these complex distribution networks with high accuracy. Therefore, an accurate, adaptive, and robust load modeling method for active distribution networks (ADNs) is in utmost need to analyze modern transmission power systems.

Recently, deep learning has been explosively developed. As deep learning has powered many research areas, the techniques for a majority of fields have been dramatically improved by using the techniques of deep learning [12]. A variety of neural networks (NNs) has been proposed in the past few years, such as deep spare autoencoder, convolutional NNs (CNNs), and recurrent NNs (RNNs). Different kinds of NN topologies are applied to different issues. For example, CNNs are more suitable for processing a grid of values, such as images, while RNNs are specially designed for processing a sequence of values. Hence, RNNs usually have the outstanding performance on speech recognition and machine translation [13]–[15]. Furthermore, the NNs show better fitting results than the polynomial regression because of the universal function fitting capability [16], [17]. In the area of electric power systems, many leading studies about dynamic state estimation [18], fault diagnose [19], [20], short-term load forecast [21], [22], dynamic security assessment [23], [24], energy management [25], [26], and attack detection [27] have been developed with the applications of deep learning.

Motivated by the similarities between power system differential algebraic equations (DAEs) and the forward calculation flows of RNNs, we try to employ RNNs from NNs for exploring a novel way to construct an equivalent model of ADNs. In this paper, a long short-term memory (LSTM) RNN is proposed to address the challenges of modeling of

Manuscript received February 12, 2018; revised August 2, 2018 and October 16, 2018; accepted November 28, 2018. This work was supported by the National Key Research and Development Program of China under Grant 2017YFB0902800. *(Corresponding author: Shaorong Wang.)*

C. Zheng and S. Wang are with the State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: kmzc1991@gmail.com; wsrwy96@vip.sina.com).

Y. Liu is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996 USA, and also with the Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA (e-mail: liu@utk.edu).

C. Liu is with the Department of Energy Technology, Aalborg University, 9220 Aalborg, Denmark (e-mail: cli@et.aau.dk).

W. Xie, C. Fang, and S. Liu are with the State Grid Shanghai Municipal Electric Power Company, Shanghai 200122, China (e-mail: xiew@sh.sgcc.com.cn; fangc02@139.com; liuxshu2012@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2018.2885219

ADNs with high accuracy. First, the adoption reasons for the proposed LSTM RNN-based equivalent model are explained, and its advantages are analyzed from the mathematical point of view. Then, in the study cases, the assumptive measurement data [e.g., phasor measurement unit (PMU) data], which are sampled from time-domain simulations, are employed to construct and test the proposed LSTM RNN-based equivalent load model of ADNs.

The key contributions of this paper are listed as follows.

1) *New Application:* It is the pioneering research to explore the feasibility of using NNs constructed by LSTM blocks and fully connected (FC) layers to build equivalent models of ADNs.
2) *Why and How:* Discussion and explanation about the reasons of adopting RNN as an equivalent model of ADNs for the study of transmission systems are originally conducted in detail. The basic rules and guidelines for designing an LSTM RNN-based equivalent model of ADNs are presented as well.
3) *New Methodology:* The proposed LSTM RNN-based equivalent model is one novel method to address the nonlinear and temporally complex challenges of ADNs, caused by the integration of renewable energies and energy storage in the distribution level of electric power grids.

The remainder of this paper is organized as follows. Section II presents the literature reviews about the load modeling methods and the application of NNs in the power system. In Section III, some important concepts of RNNs, LSTM, and backpropagation through time (BPTT) algorithm are introduced. The reason for adopting an LSTM RNN for load modeling is explained, and the basic rules and guidelines of the LSTM RNN-based equivalent model is presented. Section IV is about study cases of the LSTM RNN-based load model. The conclusion is presented in Section V.

II. Literature Review

In the past decades, load modeling approaches can be generally classified into two categories: component-based approaches [28]–[32] and measurement-based approaches [33]–[37]. Component-based modeling approaches build the mathematical functions to represent the relationship between a bus voltage/frequency and the active/reactive power consumed by loads based on individual load physical characteristics. They highly depend on the accurate information of the load composition, but the time-variant load composition cannot always be obtained in an operational power system. The measurement-based approaches employ the field data to capture the loads' characteristics. Then, the load model is constructed and the parameters are identified to represent the characteristics of loads. Therefore, measurement-based approaches can reflect the time-variant load dynamics better than the component-based approaches, which are more widely employed for power system studies in practical applications.

As there has been pressing demands for finding a new way to construct an equivalent model of aggregated loads, some researchers have proposed using NNs to build the mapping relationship between the inputs (voltage/frequency) and the outputs (active/reactive power) for load modeling in recent years [7], [38]–[43]. Chen and Mohler [7] point out the static and quasi-static models do not consider load dynamics in the voltage stability analysis. In order to make the results of voltage stability assessment more reliable, an NN model is employed taking load dynamics into account. Ku *et al.* [38] develop a feed-forward artificial NN (ANN) with a single layer of hidden nodes to describe the complex dynamic behaviors of loads. The performance of the model has been assessed by the real field data from Taiwan Power Company. Keyhani *et al.* [39] present an approach for developing the NN-based load model in the power system stability analysis. Two steps, developing an NN with simulation data and updating the NN with measured data, are involved in this method. Hiyama *et al.* [40] investigate an NN with feedback loops for modeling the loads' dynamics. The study cases show that the NN can successfully emulate the real/reactive power response. Bostanci *et al.* [41] point out that the developed NNs, designed with a delayed input and feedback connections, can represent the dynamic behavior of power system loads. Chang *et al.* [42] present an efficient method, combining the radial basis function NN and the lookup table, to predict the response of a time-variant electrical arc furnace (EAF). The proposed EAF model can be effectively used to assess waveform distortions, voltage fluctuations in power systems. In these studies, different kinds of ANNs are adopted for load modeling, but the cases with distributed generations in distribution networks have not been taken into consideration. Amy *et al.* [43] present an ANN-based model for ADNs, but the adoption of ANN and equivalent lumped passive elements to represent the ADNs may reduce the accuracy because there are no temporal connections in the ANNs.

III. Recurrent Neural Network-Based Load Model

Load modeling refers to a mathematical mapping between a bus voltage/frequency and the active/reactive power flowing into the bus [1]. This section, first, gives a brief review of some important ideas of RNNs, the structure of an LSTM block, and the BPTT algorithm. The details of an LSTM RNN can be found in [44], and the details of the BPTT algorithm can be found in [45]–[47]. Then, the motivation of adopting an LSTM RNN for modeling load dynamics is explained and discussed in detail. Finally, basic rules and guidelines for designing an LSTM RNN-based equivalent model, where LSTM blocks and FC layers are adopted as hidden layers, are introduced.

A. Some Important Concepts of Recurrent Neural Networks

RNNs are very powerful dynamic systems [12]. They are designed for dealing with sequences and have achieved huge successful performance on applications of machine translation and speech recognition, so it is often better to adopt RNNs for the tasks, which involve sequential inputs.

Some important concepts of RNNs will be presented in this section. As illustrated in Fig. 1, the input of a recurrent neuron shown in yellow is fed by the lower layer shown in blue, and

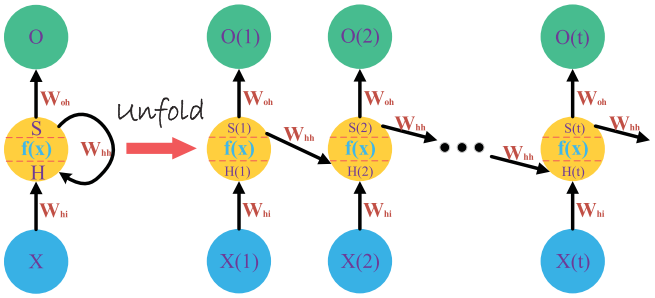

Fig. 1. Example of recurrent neuron unfolding diagram.

the output of a recurrent neuron is propagated to the higher layer shown in green. At the same time, the state $S(t)$ of a recurrent neuron is fed into this recurrent neuron at the next time step $t+1$. A block of recurrent neuron can be unrolled with finite time step [13]. It can be clearly noticed from the structure of RNN that the states of current time $t$are influenced by the information from former time steps $t-1, \cdots, 1$ in this unfolded architecture. In addition, each layer of RNNs will share its parameters in the time step of the entire sequence, this characteristic makes it possible to process variable length sequences, and scale to much longer or shorter sequences [13].

When armed with the ideas of recurrent neuron unfolding and weight sharing, the procedure of a conventional RNN is briefly introduced here. The formulations from (1) to (3) are applied to compute a sequence of hidden layer states $S(t)$ and a sequence of outputs $O(t)$

$$H(t) = W_{hi}X(t) + W_{hh}S(t-1) + b_h \tag{1}$$
$$S(t) = f(H(t)) \tag{2}$$
$$O(t) = W_{oh}S(t) + b_o \tag{3}$$

where $X(t) \in R^{n_i}$ indicates an input sequence with n_i dimensions of RNN at time step$t$, $H(t) \in R^{n_h}$ represents the linear state of a hidden layer at time step t, $S(t) \in R^{n_h}$ represents the nonlinear state from hidden layers at time step t, and $O(t) \in R^{n_o}$ indicates an output sequence with n_o dimensions of RNN at time step t. n_i, n_h, and n_o are the number of elements in input layer, hidden layer, and output layer, respectively. W_{hi}, W_{hh}, and W_{oh} are the weight matrices for corresponding layers of the standard RNN, and b_h and b_o are the bias vectors for corresponding layers of the standard RNN. $f(x)$ is a nonlinear activation function, generally, $f(x) = \tanh(x)$ or$f(x) = \text{sigmoid}(x)$. It can be noted that from the structure of the conventional RNN, all weight matrices are independent of the time step, because of the involved concept of weight sharing.

B. Long Short-Term Memory Block

As mentioned above, the traditional RNN is one of the most powerful tools to construct a temporal relationship. However, it has been proven that the traditional RNN is difficult to learn long-term dependencies, because the traditional RNN have troubles of gradient vanishing or exploding, when the gradients back propagated over some steps [48], [49]. In order to break the limits of learning long-term dependencies, the German

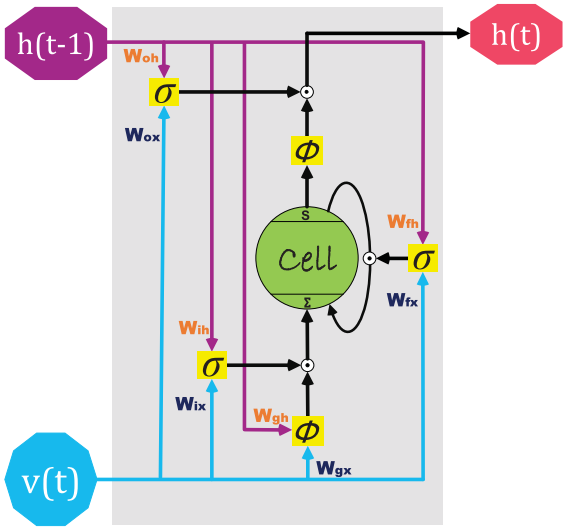

Fig. 2. Architecture of an LSTM block.

researchers S. Hochreiter and J. Schmidhuber had advanced an improved structure of recurrent neuron based on the idea of gate, which is the so-called LSTM [44], [50]. The LSTM networks have attracted enormous attention in academia and industry, because the LSTM networks are successfully applied in many fields, such as speech recognition, machine translation, and load forecast.

An LSTM block is illustrated in Fig. 2. The LSTM is constructed by the so-called memory cell state S, input node g_c, input gate i_g, output gate o_g, and forget gate f_g. The detailed description can be found in [51]. The formulations of an LSTM block are expressed in the following equations:

$$g_c(t) = \phi(W_{gx}v(t) + W_{gh}h(t-1) + b_g) \tag{4}$$
$$i_g(t) = \sigma(W_{ix}v(t) + W_{ih}h(t-1) + b_i) \tag{5}$$
$$f_g(t) = \sigma(W_{fx}v(t) + W_{fh}h(t-1) + b_f) \tag{6}$$
$$o_g(t) = \sigma(W_{ox}v(t) + W_{oh}h(t-1) + b_o) \tag{7}$$
$$s(t) = g_c(t) \odot i_g(t) + s(t-1) \odot f_g(t) \tag{8}$$
$$h(t) = \phi(s(t)) \odot o_g(t) \tag{9}$$

where $v(t)$ indicates the inputs of an LSTM block at time step t; $h(t)$ indicates the outputs of an LSTM block at time step t; W_{gx}, W_{ix}, W_{fx}, and W_{ox} represent the correlation weight matrices between the input x and the input node g_c, input gate i_g, forget gate f_g, and output gate o_g, respectively; and W_{gh}, W_{ih}, W_{fh}, and W_{oh} represent the correlation weight matrices between the last time step output h and the input node g_c, input gate i_g, forget gate f_g, and output gate o_g, respectively. Then, b_g, b_i, b_f, and b_o are biases of the input node g_c, input gate i_g, forget gate f_g, and output gate o_g; and $\odot$ stands for element-wise multiplication. $\sigma(x)$ represents an element-wise sigmoid nonlinear activation function, and $\phi(x)$ represents an element-wise hyperbolic tangent nonlinear activation function.

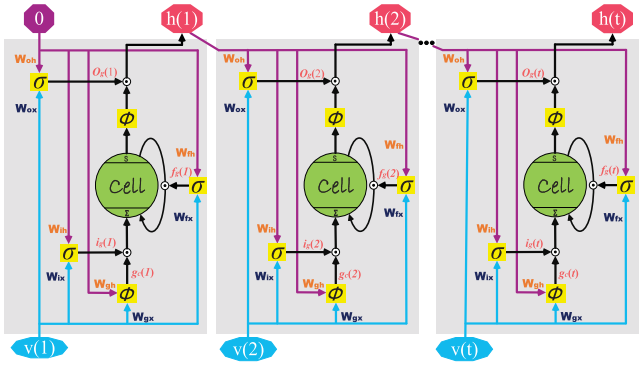

Fig. 3. Diagram of the forward part of BPTT used for the LSTM block.

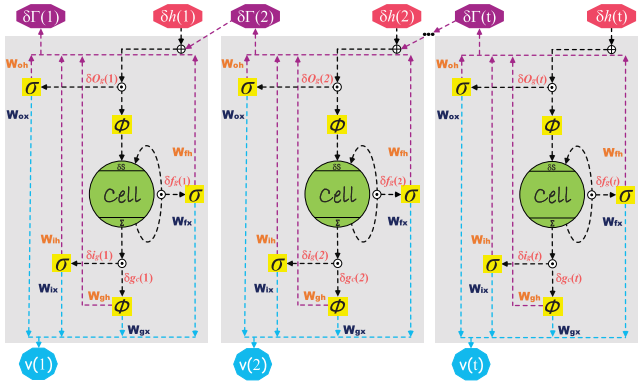

Fig. 4. Diagram of the backward part of BPTT used for the LSTM block.

C. Backpropagation Through Time Algorithm

The input of an LSTM RNN is an ordered sequence, there are some suitable algorithms, such as real-time recurrent learning, BPTT, and truncated BPTT for training of LSTM RNNs, rather than the traditional backpropagation [45]–[47]. Owing to the reason, the problems of gradient vanishing and exploding in the learning duration of LSTM are addressed, and the BPTT algorithm is adopted as a learning algorithm for training the LSTM RNNs in this paper. Generally, the BPTT algorithm consists of two steps. A forward calculation is first employed to pass through whole time steps in the unfolded calculation graph of LSTM RNN. Then, a backward calculation is applied to go back through the whole network from the last time step t_1 to the first time step t_0 in the unfolded calculation graph [52]. As described in [40], the weighted summation of the error sequences between the LSTM RNN output sequences and target sequences is a natural choice as the performance criterion, when an LSTM RNN is applied as a regression procedure. Hence, the following equation is chosen as a cost function J_{cost} in this paper:

$$J_{\text{cost}} = \frac{1}{2}\sum_{t=1}^{n}(h(t) - \hat{h}(t))^2 \tag{10}$$

where $h(t)$ denotes the output sequence of the NN at time step t, and $\hat{h}(t)$ denotes the target sequence of the NN at time step t.

The data flow in the forward and backward propagations of an LSTM block is exhibited in Figs. 3 and 4, respectively. Meanwhile, the corresponding pseudocode is illustrated in Algorithm 1 for summarizing all of the learning procedure of an LSTM block. $v(t)$ shown in Figs. 3 and 4 represents the input vector of LSTM, and $h(t)$ represents the output of LSTM. Other symbols have the same meaning with (4)–(9).

Algorithm 1 Learning Procedure of LSTM Block

Cost Funtion: $J_{cost} = \frac{1}{2}\sum_{t=1}^{N}(h(t) - \hat{h}(t))^2$
Learning Rate: η

Forward:

$J_{cost} \leftarrow 0$
for t **from** t_0 **to** t_1 **do**
  $\widetilde{g}_c(t) \leftarrow W_{gx}v(t) + W_{gh}h(t-1) + b_g$
  $g_c(t) \leftarrow \phi(\widetilde{g}_c(t))$
  $\widetilde{i}_g(t) \leftarrow W_{ix}v(t) + W_{ih}h(t-1) + b_i$
  $i_g(t) \leftarrow \sigma(\widetilde{i}_g(t))$
  $\widetilde{f}_g(t) \leftarrow W_{fx}v(t) + W_{fh}h(t-1) + b_f$
  $f_g(t) \leftarrow \sigma(\widetilde{f}_g(t))$
  $\widetilde{o}_g(t) \leftarrow W_{ox}v(t) + W_{oh}h(t-1) + b_o$
  $o_g(t) \leftarrow \sigma(\widetilde{o}_g(t))$
  $s(t) \leftarrow g_c(t) \odot i_g(t) + s(t-1) \odot f_{gg}(t)$
  $h(t) \leftarrow \phi(s(t)) \odot o_g(t)$
end for

Backward - *BPTT*:

$\delta h(t) = \frac{\partial J_{cost}}{\partial h(t)}$; $\delta f_g(t) = \frac{\partial J_{cost}}{\partial f_g(t)}$; $\delta o_g(t) = \frac{\partial J_{cost}}{\partial o_g(t)}$

$\delta s(t) = \frac{\partial J_{cost}}{\partial s(t)}$; $\delta i_g(t) = \frac{\partial J_{cost}}{\partial i_g(t)}$; $\delta g_c(t) = \frac{\partial J_{cost}}{\partial g_c(t)}$

for t **from** t_1 **down to** t_0 **do**
  $\delta\Gamma(t) \leftarrow W_{oh}\cdot\delta\widetilde{o}_g(t) + W_{fh}\cdot\delta\widetilde{f}_g(t) + W_{ih}\cdot\delta\widetilde{i}_g(t) + W_{gh}\cdot\delta\widetilde{g}_c(t)$
  $\delta h(t) \leftarrow \delta h(t) + \delta\Gamma(t+1)$
  $\delta s(t) \leftarrow \delta h(t) \odot \phi'(s(t)) \odot o_g(t)$
  $\delta\widetilde{o}_g(t) \leftarrow \delta h(t) \odot \phi(s(t)) \odot \sigma'(\widetilde{o}_g(t))$
  $\delta\widetilde{f}_g(t) \leftarrow \delta s(t) \odot s(t-1) \odot \sigma'(\widetilde{f}_g(t))$
  $\delta\widetilde{i}_g(t) \leftarrow \delta s(t) \odot g_c(t) \odot \sigma'(\widetilde{i}_g(t))$
  $\delta\widetilde{g}_c(t) \leftarrow \delta s(t) \odot i_g(t) \odot \phi'(\widetilde{g}_c(t))$
  $dW_{mx} \leftarrow dW_{mx} + Mv(t),\ m \in \{g, i, f, o\}$
  $dW_{mh} \leftarrow dW_{mh} + Mh(t-1),\ m \in \{g, i, f, o\}$
  $db_m \leftarrow db_m + M,\ m \in \{g, i, f, o\}$
  $M \in \widetilde{g}_c(t), \delta\widetilde{i}_g(t), \delta\widetilde{f}_g(t), \delta\widetilde{o}_g(t)\}$
end for

D. Reason of Choosing LSTM RNN for Load Modeling

Due to feedback connections in the hidden layer, RNNs have powerful ability to build the temporal correlations between the latest states and the previous states. On the other hand, the field data of aggregated loads in electric power systems are a sequence of sampled discrete values within a time stamp, which can be considered as an ordered sequence data in the time domain. Moreover, the disturbance occurred at time t will influence the power responds at time $t + \tau$ in power systems. Therefore, based on the mentioned sketchy content above, the RNNs would be one of the ideal selections for building an equivalent model of ADNs.

The distribution networks are regarded as an aggregated load seen from bulk transmission power delivery systems for the analysis of transmission systems in this paper. The dynamic behaviors of load elements can usually be described by DAEs. For example, the induction motor load, which is one of the most important dynamic load elements in distribution networks, can be represented by the simplified third-order DAEs [53]. Generally, the dynamic characteristics of high-order power system loads can be described in the form of the differential equations (DEs) (11) and the algebraic equations (AEs) (12) in the state space

$$\dot{x} = f(x, u, p) \tag{11}$$

$$y = g(x, u, p) \tag{12}$$

where x denotes the state vector, u denotes the input vector, y is the output vector, and p is the parameter variable. The discrete version of DEs (11) could be easily written as follows:

$$x(t) = f'(x(t-1), u(t), p). \tag{13}$$

The calculation flows of an LSTM block are enumerated in (4)–(9). The calculation flows of hidden layers formed with LSTM blocks are almost the same as the LSTM block. With the intention of clearly indicating the relationship between the inputs and the outputs of the hidden layers formed by LSTM blocks, the forward calculation procedure can be mathematically represented in the following equation:

$$h(t) = F(h(t-1), v(t), W, b) \tag{14}$$

where $v(t)$ denotes the input vector of LSTM hidden layers at time t, $h(t)$ denotes the state vector of LSTM hidden layers at time t, W denotes the weight matrix, and b denotes the bias vector. Equation (14) of the LSTM hidden layers is in the form of difference equations as well. Equation (14) for the LSTM hidden layers is almost the same as the discrete version of DEs (13) in terms of mathematical forms. Therefore, the hidden layers constructed with recurrent neurons (e.g., LSTM blocks) may have the same capability of representing dynamic systems as the mathematical representations at the state space.

As mentioned before, the part of DEs in DAEs can be represented with the hidden layers constructed via LSTM blocks. Moreover, based on the definition of load modeling and the DAEs at the state space, it is necessary to develop other hidden layers upon the proposed NNs for representing the part of AEs in DAEs. Thus, in this paper, the FC hidden layers are stacked upon the LSTM hidden layers to represent the AEs in DAEs. Then, the proposed hybrid NNs could be mathematically represented in the following equations:

$$h(t) = F(h(t-1), v(t), W_l, b_l) \tag{15}$$

$$O(t) = G(h(t), v(t), W_f, b_f) \tag{16}$$

where $v(t)$ denotes the input vector of LSTM hidden layers at time step t, $h(t)$ denotes the state vector of LSTM hidden layers at time step t, $O(t)$ denotes the output vector of FC layers at time step t, W_l denotes the weight matrix of LSTM hidden layers, b_l denotes the bias vector of LSTM hidden layers, W_f denotes the weight matrix of FC hidden layers, and b_f denotes the bias vector of FC hidden layers. The mathematical formulations of the proposed hybrid NN are similar to the discrete version of DAEs.

Due to these similarities, it has demonstrated that the hidden layers constructed with LSTM blocks can be employed to represent the states of ADNs (the part of DEs), and the FC hidden layers could be employed for combining the states learned with LSTM hidden layers to acquire the outputs—real and reactive power responses (the part of AEs). In this proposed hybrid NN, the states of ADNs are developed by learning from the training sets, rather than manually defined. Therefore, motivated by these factors, the LSTM-based hybrid NN (the so-called LSTM RNN) is employed for representing the aggregated dynamic behaviors of ADNs in electric power systems.

E. Method of Constructing an LSTM RNN as an Equivalent Load Model of ADNs

Armed with foundational knowledge about architecture, feed-forward calculation procedure, learning, and optimizing technique of NNs, all the things are ready for designing a suitable NN to address the special issue—building an equivalent model of ADNs.

As mentioned above, the basic rules of designing an LSTM RNN are: 1) employing the hidden layers constructed with LSTM blocks to represent DEs of DAEs and adopting FC hidden layers to represent AEs of DAEs and 2) the FC layers are stacked upon the LSTM hidden layers (i.e., the outputs of the hidden layer constituted by LSTM blocks are fed into the FC hidden layers).

Moreover, the number of LSTM hidden layers and FC layers are influenced by many factors, such as the scale of ADNs, types of load elements, and penetration of renewable energies. Due to this reason, the number of LSTM hidden layers and FC layers are usually determined by experiments considering the performance and computational burden of LSTM RNNs. There are basic guidelines to determine the number of LSTM hidden layers and FC layers: 1) based on aforementioned rules, there must be at least one LSTM hidden layer and one FC layer in the hidden layers of LSTM RNN and 2) due to the fact that LSTM RNNs could represent more complex ADNs via increasing LSTM hidden layers or FC hidden layers: a) the number of LSTM hidden layer should be increased if there are various types or large scales of dynamic elements (motor machines, electrical vehicles, and renewable energies) in ADNs, because the LSTM hidden layers are adopted as DEs and b) the number of FC hidden layers should be increased if there are high percentage or various kinds of static elements (lightings and heaters), because the FC layers are used as AEs.

In the light of the definition of the load modeling in electrical power systems, typically, the bus voltage within an N-step time window is selected as the inputs, and the corresponding real and reactive powers flowing into the bus are selected as the outputs, when an LSTM RNN is adopted for constructing an equivalent model of ADNs in electric power systems.

Furthermore, some key points should be paid attention when employing an LSTM RNN for load modeling. It is

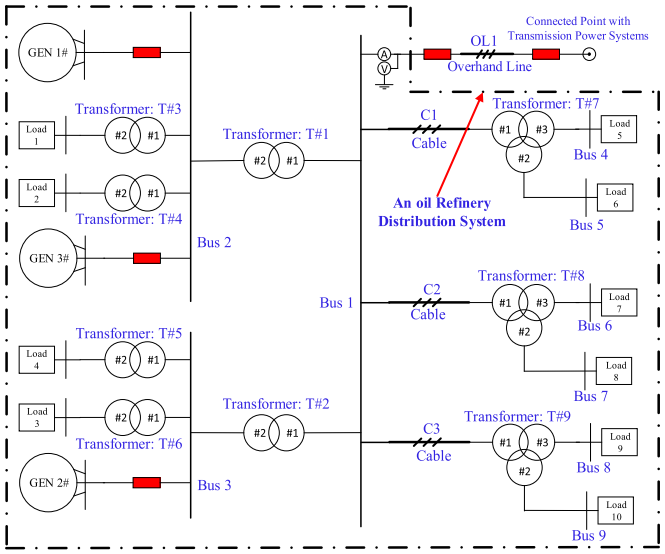

Fig. 5. Framework of a real oil refinery distribution system.

well known that the nonlinear activation function (the sigmoid and the hyperbolic tangent activation function) employed by the LSTM blocks and FC hidden layers will arrive into the saturation area, if the input values are beyond the limits of the linear. Because of this characteristic of the activation functions, it is critical to scale all inputs within the range of (0, 1). Thus, the preprocessing step (normalization) should be adopted before the N-step time window bus voltages enter the input layer of the proposed LSTM RNN. While the targets of the proposed LSTM RNN need not to be normalized, because there is no existing nonlinear activation function in the output layer.

IV. Study Cases: Evaluating the Performance of the LSTM RNN-Based Equivalent Model

In this section, an electric power testing system containing detailed ADNs is first introduced. Based on the basic rules and guidelines for designing the LSTM RNN-based equivalent model, five different LSTM RNNs are designed. Then, the LSTM RNN built with one LSTM hidden layer and two FC hidden layers is selected to represent the testing system by comparing the experiment results. Afterward, the performance of the LSTM RNN will be comprehensively evaluated with a variety of assumptive measurement data (e.g., PMU data), which are sampled from the time-domain simulation of the testing system. Finally, the LSTM RNN-based equivalent model is compared with ANN-, conventional RNN-, support vector regression (SVR)-, and nonlinear autoregression with exogenous (NARX)-based equivalent model. Meanwhile, in the study case, the frameworks of the LSTM RNN-based equivalent model are built by using TensorFlow [54], and the LSTM RNN are trained and tested with a personal computer equipped with Intel Core i7-4790 CPU, 3.6-GHz processor, and 16 GB of RAM.

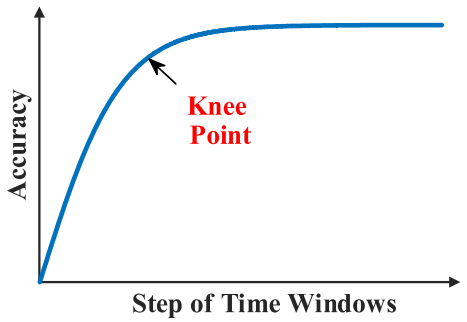

Fig. 6. Trajectory about final average accuracy and step of the time window.

A. Brief Introduction of Testing System

The testing system contains two parts, namely, IEEE 39-Bus New England system [55], and a detailed ADN. The New England system is employed as the transmission system, and the ADN is adopted as the distribution system. Besides, the detailed distribution network connects with the Bus-39 of the New England test system through an electrical power transmission line to replace a part of the original loads at Bus-39. The details of the New England system is presented in [55]. The parameters and topology of the distribution network illustrated in the Appendix and Fig. 5, respectively, are from a realistic oil refinery. In addition, the entire testing system is constituted in detail by using the PSCAD software. In the detailed distribution network, the static loads, dynamic loads, and electric generators are separately modeled by the fixed load model, squirrel cage induction machine model, and synchronous machine model, which are provided by the PSCAD software, respectively.

The frequency of this testing system is 50 Hz. In this study case, the time-domain simulation step is 50 μs. At the same time, based on the basic requirement of the Nyquist–Shannon sampling theorem, the time-domain simulation results are continuously sampled at a sampling data rate of 1 ms per point for producing the correlated field data, which are acquired by measurement devices (e.g., PMUs).

B. Constructing LSTM RNN as an Equivalent Model of ADNs

Generally, the hyperparameters tuning (i.e., the number of hidden layers and the number of neuron per layer) is a tough and time-consuming task, which is also faced in most of the applications of deep learning. The trajectory of final average accuracy and the length of input time step window are illustrated in Fig. 6 for most of the cases. The accuracy will enter the saturation area when the step of the time window exceeds the "knee point." However, the longer the step of time window, the more computation time may be taken. Due to a tradeoff between these factors, the knee point of Fig. 6 is usually selected as the step of time window. Thus, the bus voltages within five-step time window are selected as inputs of LSTM RNN (i.e., the number of neurons in the input layer is five) in this study case. Moreover, according to the objective of load modeling, the real power response or reactive power

TABLE I
EXPERIMENT RESULTS OF FIVE DIFFERENT LSTM RNNS

Type	Real power		Reactive power		Average Time Consuming
	Error of Training	Error of Testing	Error of Training	Error of Testing	
Model I	0.053	0.047	0.113	0.083	5h
Model II	0.043	0.039	0.111	0.079	5.3h
Model III	0.048	0.045	0.111	0.075	6h
Model IV	0.032	0.058	0.085	0.097	6.5h
Model V	0.028	0.068	0.080	0.105	6.7h

response is selected as the output of LSTM RNN, so the number of output layers is one in this study case.

Based on the experience from lots of experiments, three or four hidden layers of LSTM RNN are enough for constructing an equivalent model of the ADN shown in Fig. 5. Furthermore, according to the basic rules of designing and guidelines listed in Section III, five different LSTM RNNs have been designed to be selected by the experimental method.

1) *Model I:* one input layer, one LSTM hidden layer, two FC hidden layers, and one output layer.
2) *Model II:* one input layer, two LSTM hidden layers, one FC hidden layer, and one output layer.
3) *Model III:* one input layer, one LSTM hidden layer, three FC hidden layers, and one output layer.
4) *Model IV:* one input layer, two LSTM hidden layers, two FC hidden layers, and one output layer.
5) *Model V:* one input layer, three LSTM hidden layers, one FC hidden layer, and one output layer.

These five models are tested with a part of the time-domain simulation results of the study cases. The testing results illustrated in Table I demonstrate as follows.

1) Overfitting occurred in Model IV and Model V, because the average testing errors of these two models are increased, compared with the other models.
2) The average testing errors of Model I, Model II, and Model III are almost the same.
3) The more the number of the hidden layer, the more computation time is taken.
4) When the number of the hidden layers is the same, the more the number of LSTM hidden layer, the more computation time is taken.

With considering the tradeoff between the computational burden and the performance (i.e., the average training time of Model I is shortest, and the average testing error of Model I is acceptable), Model I, illustrated in Fig. 7, is employed for constructing an equivalent model of the ADNs in this study case. Besides, the LSTM hidden layer contains 100 LSTM blocks, and two FC hidden layers contain 150 and 80 neurons, respectively.

Some optimization tools, such as RMSProp optimizer [56], Adagrad optimizer [57], Adadelta optimizer [58], and Adam optimizer [59], are principally tested with the employed LSTM RNN. The testing results are illustrated in Fig. 8. It is observed that Adam optimizer performs better than the other tools, which is hence adopted for the training of an LSTM RNN model in this paper. Meanwhile, the default hyperparameters

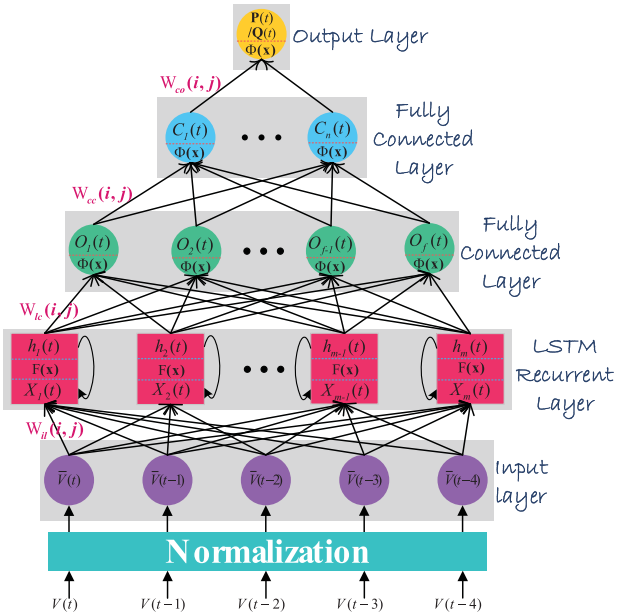

Fig. 7. RNN-based load model framework. $\phi(x)$: hyperbolic tangent activation function; $F(x)$: mapping relationship between the inputs and outputs of LSTM blocks; $\bar{V}(t)$: variables of input layers; $X_i(t)$: inputs of the ith LSTM block; $h_i(t)$: outputs of the ith LSTM block; $O_i(t)$: variables of ith neurons at the first FC layer; $C_i(t)$: variables of ith neurons at the second FC layer; and $P(t)/Q(t)$: variable of output layer.

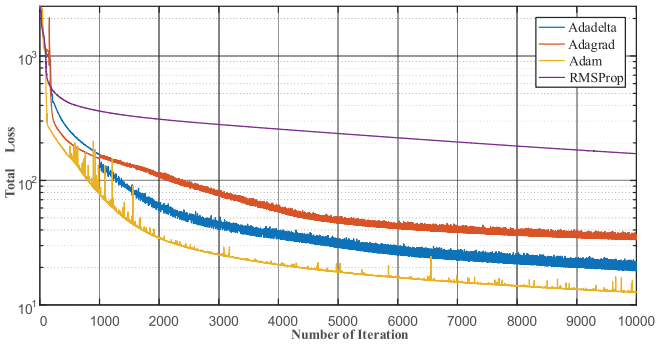

Fig. 8. Comparison results of the optimization candidates.

($\mu = 0.9$, $\nu = 0.999$, and $\varepsilon = 10^{-8}$), recommended in [59], are selected as the hyperparameters of the Adam optimizer.

C. Performance Evaluation of the Proposed LSTM RNN-Based Equivalent Model

Due to the factors that three-phase short-circuit fault in the transmission system is the most serious disturbance, and the dynamic behaviors of aggregated loads are essential for the stability analysis in power systems [37], the three-phase short-circuit faults are then adopted in the study case to test the dynamic behavior of ADNs.

The time-domain simulation results contain 36 pairs of real and reactive power responses caused by different disturbances. Besides, 25 pairs of real and reactive power responses are from the cases, where the disturbances (i.e., three-phase short-circuit fault through grounding resistance) occur at the point of connection in the New England transmission

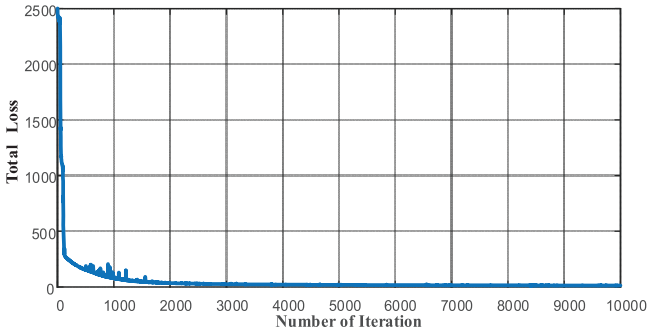

Fig. 9. Cost function for the whole learning procedure.

system, and the severity of the disturbances is different by adjusting the grounding resistance. Another 11 pairs of real and reactive power responses are from the cases, where the disturbances occur at different places of the New England transmission system. Then, assumptive field data sampled from time-domain simulations are separated into two independent sets (i.e., a training set and a testing set) in this study case. Generally, the ratio of the simulation results contained in the training set to entire simulation results is 75% (i.e., 27 simulation results), and the ratio for the testing set is the remaining 25% (i.e., nine simulation results).

The root-mean-square error (RMSE) is adopted as an index to evaluate the matching performance between the outputs by the LSTM RNN-based equivalent model and response for the entire time-domain simulation. Meanwhile, relative error per point (REP) is employed as another index to describe the maximum error generated within the entire dynamic interval. The RMSE and REP are calculated by the following equations:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{17}$$

$$\text{REP}_i = \frac{|y_i - \hat{y}_i|}{y_{sd}} \times 100\% \tag{18}$$

where y_i is the detailed time-domain simulation value at time step i, $\hat{y}_i$ is the output of the proposed LSTM RNN-based equivalent model at time step i, y_{sd} is the simulation value at steady state before disturbance, and N denotes the total number of sampled data during entire time-domain simulation.

In order to construct an equivalent model of ADNs using the designed LSTM RNN, the training set is employed to identify weights and biases of the LSTM RNN. The total cost of the LSTM RNN during the learning procedure reduces rapidly and stabilizes to a small value, as illustrated in Fig. 9. It indicates that the LSTM RNN-based load model has the capability of reproducing the dynamic behaviors of the ADNs.

The generalization capacity of the trained LSTM RNN-based equivalent model is essential for real applications. The testing set is adopted for evaluating the generalization capability of the trained LSTM RNN-based equivalent model. After that, the RMSE of real and reactive power responses in the training set and the testing set are listed in Table II. It demonstrates as follows.

TABLE II
RMSE OF THE TRAINED LSTM RNN-BASED EQUIVALENT MODEL

Type	Num	Grounding Resistance/Ω or fault location	RMSE of Real Power	RMSE of Reactive Power
Training Set	1	1	0.13711	0.11722
	2	8*	0.10684	0.08772
	3	10	0.08621	0.08621
	4	13	0.07917	0.08045
	5	17	0.06766	0.06954
	6	20	0.06130	0.05386
	7	22	0.05647	0.04683
	8	25	0.04939	0.03947
	9	27	0.04567	0.03617
	10	30	0.04183	0.03276
	11	33	0.03956	0.03028
	12	38	0.03739	0.02862
	13	40	0.03676	0.02853
	14	43	0.03597	0.02843
	15	45	0.03556	0.02831
	16	48	0.03516	0.02806
	17	50	0.03509	0.02791
	18	60	0.03779	0.02835
	19	300	0.03441	0.05949
	20	Bus2	0.04537	0.03826
	21	Bus8	0.03348	0.02949
	22	Bus9	0.03128	0.03523
	23	Bus39	0.09569	0.10568
	24	Line1-2	0.04261	0.03586
	25	Line8-9	0.03216	0.03325
	26	Line1-39/L*	0.04213	0.02851
	27	Line1-39/S	0.04020	0.03154
Testing Set	1	5*	0.53480	0.11840
	2	15	0.07286	0.07719
	3	35	0.03855	0.02923
	4	55	0.14842	0.34095
	5	100	0.05149	0.03457
	6	200	0.04612	0.05108
	7	Bus 1*	0.05175	0.04314
	8	Line1-39/M	0.03856	0.03050
	9	Line9-39	0.03452	0.02569

* indicates this case is presented with figure.
L/M/S indicate different distances between the bus 39 and the place where disturbance occurs. L donates the distance is 20 km, M donates the distance is 50 km, and S donates the distance is 130km.

1) The maximum RMSE of the real power response is 0.137113 in the training set and 0.534795 in the testing set.
2) The maximum RMSE of the reactive power response is 0.117219 in the training set and 0.340950 in the testing set.

On the other hand, these RMSEs, especially the RMSEs of testing set, reveal that the LSTM RNN-based equivalent model works well to represent the dynamic behaviors of the ADNs, because the RMSEs of real and reactive responses are small in both training set and testing set.

To visualize the comparison results between the outputs of the proposed LSTM RNN-based equivalent model and the simulation results with detailed model (i.e., benchmark) in the training set, two representative instances called TrC-1 and TrC-2, which are marked with asterisk in Table II, are illustrated in the form of dynamic response curves shown

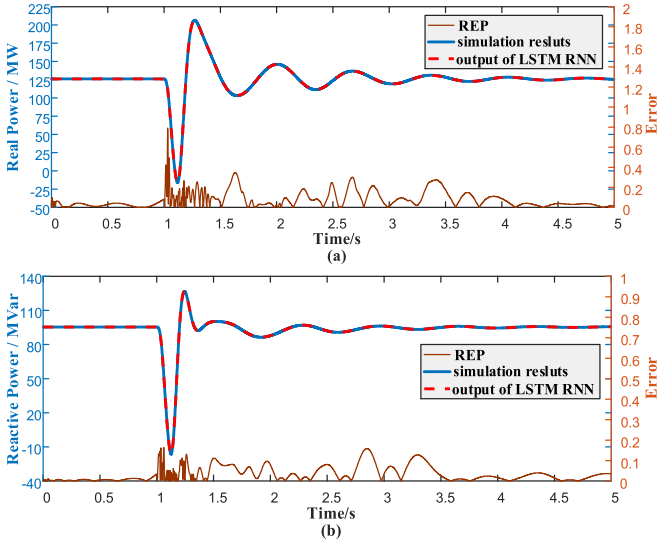

Fig. 10. Disturbance (three-phase short-circuit fault through 8-Ω grounding resistance) occurs at the point of connection in the New England transmission system. (a) Simulation result of real power and real power responses of the proposed LSTM RNN-based load model. (b) Simulation result of reactive power and reactive power responses of the proposed LSTM RNN-based load model.

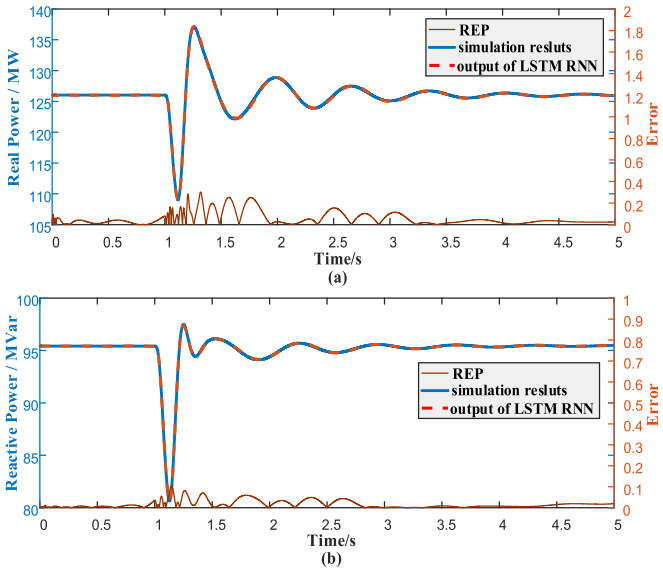

Fig. 11. Disturbance (three-phase short-circuit fault) occurs at line1-39/L of the New England transmission system. (a) Simulation result of real power and real power responses of the proposed LSTM RNN-based load model. (b) Simulation result of reactive power and reactive power responses of the proposed LSTM RNN-based load model.

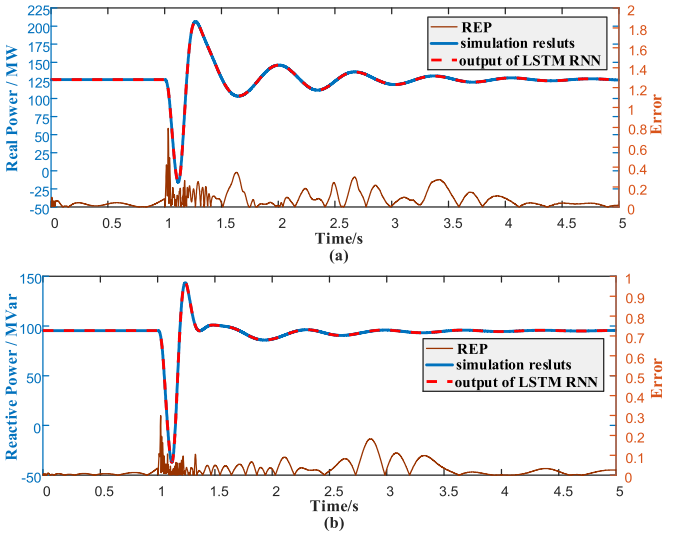

Fig. 12. Disturbance (three-phase short-circuit fault through 5-Ω grounding resistance) occurs at the point of connection in the New England transmission system. (a) Simulation result of real power and real power responses of the proposed LSTM RNN-based load model. (b) Simulation result of reactive power and reactive power responses of the proposed LSTM RNN-based load model.

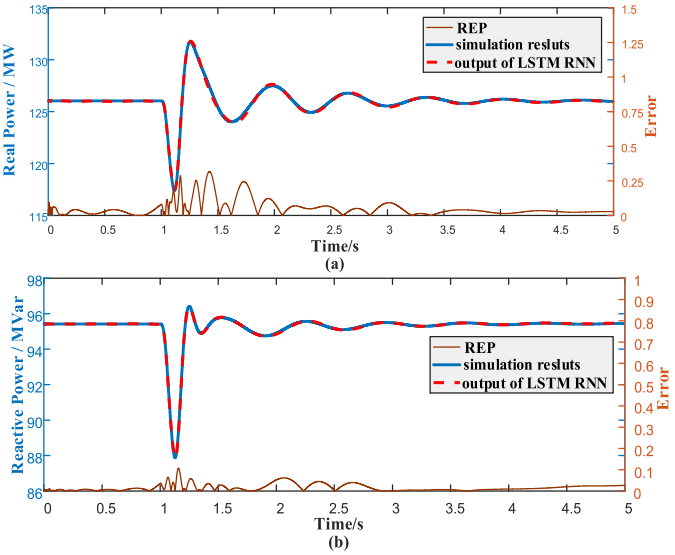

Fig. 13. Disturbance (three-phase short-circuit fault) occurs at Bus 1 of the New England transmission system. (a) Simulation result of real power and real power responses of the proposed LSTM RNN-based load model. (b) Simulation result of reactive power and reactive power responses of the proposed LSTM RNN-based load model.

in Figs. 10 and 11. As shown in Fig. 10, the peak error of TrC-1 is 0.790% for the real power response and 0.164% for the reactive power response. Likewise, the error information of TrC-2, the peak error is 0.303% for the real power response and 0.104% for the reactive power response, is shown in Fig. 11.

Figs. 10 and 11 present that the output curves (i.e., the red dashed curves) of the proposed LSTM RNN-based equivalent model match the benchmark curves (i.e., the blue curves) well within the interval of overshoot, oscillations, and steady state. Hence, the results confirm that the proposed LSTM RNN-based equivalent model can accurately estimate the dynamic behaviors (i.e., real and reactive power responses) of aggregated loads when the disturbances occur in transmission systems.

The same visualization procedures of comparison results are conducted in the testing set as well. Two representative instances, which are marked with an asterisk shown in Table II and named as TeC-1 and TeC-2, are exhibited in Figs. 12 and 13. The peak error of TeC-1 acquired from Fig. 12 is 1.225% for the real power response and 0.299% for

TABLE III
RMSE OF THE LSTM RNN-, ANN-, RNN-, SVR-, AND NARX-BASED EQUIVALENT MODEL

Num	Disturbance	RMSE of Real Power					RMSE of Reactive Power				
		LSTM	RNN	ANN	SVR	NARX	LSTM	RNN	ANN	SVR	NARX
1	5	0.535	0.865	0.951	0.732	5.884	0.118	0.391	0.581	0.426	1.151
2	15	0.073	0.262	0.495	0.543	5.772	0.077	0.331	0.572	0.501	1.027
3	35	0.039	0.258	0.501	0.485	5.364	0.029	0.359	0.498	0.475	1.069
4	55	0.148	0.295	0.486	0.512	5.459	0.341	0.385	0.562	0.408	0.995
5	100	0.052	0.186	0.521	0.487	5.648	0.035	0.326	0.573	0.421	1.047
6	200	0.046	0.278	0.492	0.463	5.523	0.051	0.341	0.542	0.435	1.035
7	Bus 1	0.052	0.193	0.476	0.491	5.641	0.043	0.321	0.595	0.459	1.075
8	Line1-39/M	0.039	0.212	0.483	0.503	5.632	0.031	0.315	0.485	0.433	1.059
9	Line9-39	0.035	0.236	0.478	0.480	5.561	0.026	0.324	0.574	0.464	1.068

the reactive power response. Similarly, as illustrated in Fig. 13, the peak error of TeC-2 is 0.317% for the real power response and 0.107% for the reactive power response. Figs. 12 and 13 present that the output curves (i.e., the red dashed curves) of the proposed LSTM RNN-based equivalent model fit the benchmark curves (i.e., the blue curves) well within the interval of overshoot, oscillations, and steady state although the disturbances are not met in the training set before. Thus, the results validate that the proposed LSTM RNN-based equivalent model can accurately represent the dynamic behaviors (i.e., real and reactive power responses) of aggregated loads, when the disturbances, which are not met in the training set before, take place in transmission systems. Furthermore, the comparison results highlight the generalization and robustness abilities of the proposed LSTM RNN-based equivalent model for representing the dynamic characteristics of ADNs.

It is clearly observed that the real and reactive power responses of aggregated loads can be regarded as two stages after voltage disturbance, as shown in the visualized figures. Besides, the first stage contains overshoot and ripples (i.e., oscillations), and the steady state is gradually set up at the second stage. Based on the results above, it is concluded that the proposed LSTM RNN-based equivalent model is robust enough for representing the dynamic behavior of ADNs in power systems.

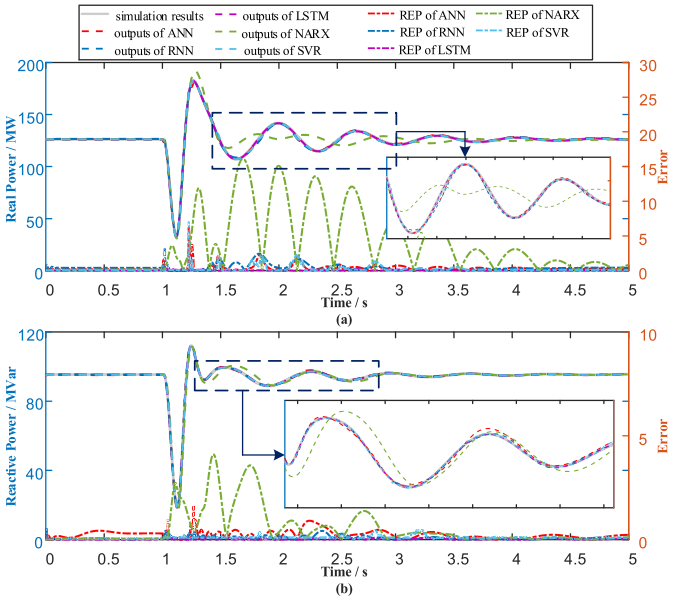

Fig. 14. Disturbance (three-phase short-circuit fault through 15-Ω grounding resistance) occurs at the point connected with the New England transmission system. (a) Simulation result of real power and real power responses of the proposed LSTM RNN-, ANN-, RNN-, SVR-, and NARX-based load model. (b) Simulation result of reactive power and reactive power responses of the proposed LSTM RNN-, ANN-, RNN-, SVR-, and NARX-based load model.

D. Performance Comparison

In this section, the performance of the proposed LSTM RNN-based equivalent model is compared with the performance of ANN-based load model [43], conventional RNN-based model [60], SVR-based model [61], and NARX-based model [62]. In order to fairly compare the performances, the ANN and conventional RNN are constructed with three hidden layers, which are same as the proposed model.

To visualize the comparison results, a representative result of these five different models is exhibited in Fig. 14. In Fig. 14, the gray curves are dynamic behaviors (i.e., real and reactive power response) of ADNs with the time-domain detailed model, and regarded as the benchmark. The purple, blue, red, green, and light blue dashed curves are dynamic behaviors of ADNs with the LSTM RNN-, conventional RNN-, ANN-, NARX-, and SVR-based equivalent model, respectively. Fig. 14 indicates the following:

1) The NARX-based equivalent model performs worst among these five methods in the interval of oscillations (ripples).
2) The SVR-based equivalent model and the ANN-based equivalent model mismatch the targets at the first stage (i.e., the magnitude of overshoot and frequency of ripples), but the SVR-based equivalent model outperforms the ANN-based equivalent model.
3) The conventional RNN-based equivalent model matches the targets better than the SVR-based equivalent model in most of testing cases.
4) The LSTM RNN-based equivalent model reaches the best fit among these five models.
5) The maximum REP of the NARX-based equivalent model is 12.793% for real power and 4.298% for reactive power, which are the highest.
6) Maximum REP of the LSTM RNN equivalent-based model is 0.428% for real power and 0.113% for reactive power, which are the lowest.

TABLE IV

COMPUTATIONAL CONSUMING AND AVERAGE TESTING RMSE OF THE LSTM RNN-, ANN-, RNN-, SVR-, AND NARX-BASED EQUIVALENT MODEL

Model	Average Training Time	Average Testing RMSE	
		Real Power	Reactive Power
ANN Based Model	3.6 h	0.543	0.553
RNN Based Model	4.8 h	0.309	0.344
SVR Based Model	3.3 h	0.522	0.447
NARX Based Model	20 min	5.609	1.058
LSTM Based Model	5 h	0.113	0.083

TABLE V

PARAMETERS OF TRANSMISSION LINE

Parameter		Length (km)	Positive Sequence(Ω)		Zero Sequence(Ω)	
Symbol	Description		R1	X1	R0	X0
OL1	Overhead Line	7.85	1.036	3.03	3.109	9.09
C1	Cable	1.7	0.132	0.368	0.453	0.98
C2	Cable	1.7	0.132	0.368	0.453	0.98
C3	Cable	1.0	0.0779	0.217	0.2666	0.98

TABLE VI

PARAMETERS OF TWO-WINDING TRANSFORMER

Parameter		Value	
Symbol	Description	T#1/2	T#3/4/5/6
	Connection	yd-11	yd-11
S_{N1}	Nominal Power	63MVA	40MVA
U_{N1}/U_{N2}	Nominal Voltage	110/38.5kV	38.5/6.3kV
U_k (%)	Short-circuit voltage ratio	10.35%	7.66%
ΔP_s	Short-circuit Loss	239.6kW	139.2kW
ΔP_0	No-load loss	36.9kW	29.7kW
ΔI_0	No-load current	0.08%	0.16%

The RMSEs of the trained LSTM RNN-, conventional RNN-, ANN-, SVR-, and NARX-based equivalent model are compared in Table III for each testing cases, and the average training time and RMSE of real power and reactive power in the testing set are listed in Table IV. Tables III and IV indicate the following.

1) Overall, the RMSEs of these models are small except the RMSE of the NARX-based model.
2) The training computation of the NARX-based model is efficient, but its performance is worst among these five models.
3) The computational consumption (i.e., average training time) for training of RNN- and LSTM-based model is almost the same, but the LSTM-based model outperforms the RNN-based model.
4) Compared with the RMSEs of the other ones, the RMSE of the trained LSTM RNN-based equivalent model is the lowest in the testing set.

Therefore, based on the comparison results shown in Fig. 14 and Tables III and IV, it is confirmed that the LSTM RNN-based equivalent model outperforms the other equivalent models for representing the dynamic characteristics of ADNs in transmission power systems.

V. CONCLUSION

An LSTM RNN-based equivalent model is proposed for representing the dynamic behaviors of ADNs in this paper.

TABLE VII

PARAMETER OF THREE-WINDING TRANSFORMER

Parameter		Value		
Symbol	Description	T#7	T#8	T#9
	Connection	ydd-11	ydd-11	ydd-11
S_N	Nominal Power	20/10/10MVA	40/20/20MVA	50/25/25MVA
U_N	Nominal Voltage	110/6.3/6.3kV	110/6.3/6.3kV	110/6.3/6.3kV
$U_{s(1-2)}$	Short-circuit voltage ratio between primary and secondary windings	13.433%	16.635%	14.873%
$U_{s(2-3)}$	Short-circuit voltage ratio between secondary and tertiary windings	25.92%	66.5%	59.5%
$U_{s(1-3)}$	Short-circuit voltage ratio between primary and tertiary windings	13.433%	16.635%	14.876%
ΔP_s	Short-circuit loss	86.5kW	172.54kW	209.12kW
ΔP_0	No-load loss	26.2kW	37.59kW	35.52kW
ΔI_0	No-load current	0.67%	0.25%	0.74%

TABLE VIII

MACHINE PARAMETERS OF SYNCHRONOUS GENERATOR

Parameter		Units	Value	
Symbol	Description		Gen 1#	Gen 2#/3#
S_N	Nominal Power	MVA	75	50
cosϕ	Nominal Power Factor		0.8	0.8
U_N	Nominal line-to-line voltage	kV	6.3	6.3
$2H = T_J$	Inertia Constant	s	2.432	2.432
R_a	Stator Resistance	p.u.	0.00358	0.00201
X_d	Synchronous reactance (d-axis)	p.u.	2.095	2.2377
X'_d	Transient reactance (d-axis)	p.u.	0.269	0.2333
T'_{d0}	Open-circuit transient time constant (d-axis)	s	6.189	6.8981
X''_d	Subtransient reactance (d-axis)	p.u.	0.215	0.1576
T''_{d0}	Open-circuit subtransient time constant (d-axis)	s	0.023	0.2051
X_q	Synchronous reactance (q-axis)	p.u.	2.0	2.20
X'_q	Transient reactance (q-axis)	p.u.	0.378	0.378
T'_{q0}	Open-circuit transient time constant (q-axis)	s	0.527	0.527
X''_q	Subtransient reactance (q-axis)	p.u.	0.215	0.215
T''_{q0}	Open-circuit subtransient time constant (q-axis)	s	0.02454	0.02454

A modified New England system is adopted to evaluate the performance of the proposed LSTM RNN-based equivalent model. The proposed LSTM RNN-based equivalent model can

TABLE IX
EXCITATION PARAMETERS OF SYNCHRONOUS GENERATOR

Parameter		Units	Value	
Symbol	Description		Gen 1#	Gen 2#/3#
T_r	Loss-pass filter time constant	s	0.02	0.02
K_A	Regulator gain	p.u.	400	400
T_A	Regulator time constant	p.u.	0.02	0.02
V_{rmax}	Regulator Output upper limit	p.u.	6.03	6.03
V_{rmin}	Regulator Output lower limit	p.u.	-6.03	-6.03
K_F	Damping filter gain	p.u.	0.03	0.03
T_F	Damping filter time constant	s	1.0	1.0

accurately represent the dynamic behaviors of ADNs in both the training set and testing set. It is also validated that the proposed LSTM RNN-based equivalent model has the generalization capability of representing the dynamic behaviors (real and reactive power responses) of ADNs in electric power systems. Furthermore, compared with the performance of ANN-, conventional RNN-, SVR-, and NARX-based equivalent models, the LSTM RNN-based equivalent model outperforms the other equivalent models in the study cases. Compared with other load modeling methods, the major advantages of LSTM RNN-based equivalent model are listed as follows.

1) The proposed LSTM RNN-based equivalent model can build the states of ADNs via learning. However, the states of DAEs of ADNs need to be built by experts and engineers according to the knowledge of electrical engineering.
2) The proposed LSTM RNN-based equivalent model can represent the dynamic behaviors of ADNs. However, weaknesses of conventional methods, e.g., low accuracy, mismatching the frequency of ripples, and failing to capture dynamic behaviors, may degrade the performance of the equivalent model of ADNs on reproducing.

Therefore, the proposed LSTM RNN-based equivalent model can be considered as one of the useful tools to represent the dynamic behaviors of ADNs. Moreover, the proposed LSTM RNN-based equivalent model can be also adopted for power stability control and voltage stability assessment in the analysis of electric power systems.

Future works include further exploring an easier way to address noise problems of realistic field measurement data in practical industry and accelerating the learning speed, so that the proposed LSTM RNN-based model can be used in online.

APPENDIX

The 10-machine 39-Bus (New England) system's parameters are listed in [55]. The parameters of the oil refinery distribution system are listed in Tables V–IX.

ACKNOWLEDGMENT

The authors would like to thank Google, Mountain View, CA, USA, for the open source of the TensorFlow that is used in this paper.

REFERENCES

[1] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994, pp. 45–138.
[2] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load modeling—A review," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5986–5999, Nov. 2018.
[3] W. W. Price *et al.*, "Load representation for dynamic performance analysis (of power systems)," *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 472–482, 1993.
[4] IEEE Task Force on Load Representation for Dynamic Performance, "Standard load models for power flow and dynamic performance simulation," *IEEE Trans. Power Syst.*, vol. 10, no. 3, pp. 1302–1313, Aug. 1995.
[5] J. V. Milanovic and I. A. Hiskens, "Effects of load dynamics on power system damping," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 1022–1028, May 1995.
[6] C. Taylor, N. Balu, and D. Maratukulam, *Power System Voltage Stability* (EPRI Series). New York, NY, USA: McGraw-Hill, 1994.
[7] D. Chen and R. R. Mohler, "Neural-network-based load modeling and its use in voltage stability analysis," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 4, pp. 460–470, Jul. 2003.
[8] W.-S. Kao, C. Lin, C.-T. Huang, Y.-T. Chen, and C.-Y. Chiou, "Comparison of simulated power system dynamics applying various load models with actual recorded data," *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp. 248–254, Feb. 1994.
[9] W. Yao, L. Jiang, J. Wen, Q. Wu, and S. Cheng, "Wide-area damping controller for power system interarea oscillations: A networked predictive control approach," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 1, pp. 27–36, Jan. 2015.
[10] A. Rouhani and A. Abur, "Real-time dynamic parameter estimation for an exponential dynamic load model," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1530–1536, May 2016.
[11] D. N. Kosterev, C. W. Taylor, and W. A. Mittelstadt, "Model validation for the August 10, 1996 WSCC system outage," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 967–979, Aug. 1999.
[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[14] Y. Zhang, P. Li, Y. Jin, and Y. Choe, "A digital liquid state machine with biologically inspired learning and its application to speech recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2635–2649, Nov. 2015.
[15] J.-T. Chien and Y.-C. Ku, "Bayesian recurrent neural network for language modeling," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 361–374, Feb. 2016.
[16] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
[17] T. Chen and H. Chen, "Approximations of continuous functionals by neural networks with application to dynamic systems," *IEEE Trans. Neural Netw.*, vol. 4, no. 6, pp. 910–918, Nov. 1993.
[18] L. Hu, Z. Wang, and X. Liu, "Dynamic state estimation of power systems with quantization effects: A recursive filter approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1604–1614, Aug. 2016.
[19] C. Dong, Y. Zhao, and Q. Zhang, "Assessing the influence of an individual event in complex fault spreading network based on dynamic uncertain causality graph," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1615–1630, Aug. 2016.
[20] B. Y. Vyas, B. Das, and R. P. Maheshwari, "Improved fault classification in series compensated transmission line: Comparative evaluation of chebyshev neural network training algorithms," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1631–1642, Aug. 2016.
[21] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.
[22] A. Almalaq and G. Edwards, "A review of deep learning methods applied on load forecasting," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 511–516.
[23] Y. Xu *et al.*, "Assessing short-term voltage stability of electric power systems by a hierarchical intelligent system," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1686–1696, Aug. 2016.
[24] H. Sun *et al.*, "Automatic learning of fine operating rules for online power system security control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1708–1719, Aug. 2016.
[25] G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam, and A. Ahmadi, "Dynamic energy management system for a smart microgrid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1643–1656, Aug. 2016.

[26] C. Napoli, G. Pappalardo, G. M. Tina, and E. Tramontana, "Cooperative strategy for optimal management of smart grids by wavelet RNNs and cloud computing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1672–1685, Aug. 2016.

[27] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1773–1786, Aug. 2016.

[28] *WECC MVWG Load Model Report Version 1.0*, Western Electricity Coordinating Council (WECC)-Modeling and Validation Work Group (MVWG), Jun. 2012. [Online]. Available: https://www.wecc.biz/Reliability/WECC%20MVWG%20Load%20Model%20Report%20ver%201%200.pdf

[29] A. Ellis, "Advanced load modeling," Electr. Power Res. Inst., Palo Alto, CA, USA, Tech. Rep. 1007318, Sep. 2002.

[30] L. Dzafic, M. Glavic, and S. Tesnjak, "A component-based power system model-driven architecture," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 2109–2110, Nov. 2004.

[31] A. Gaikwad, P. Markham, and P. Pourbeik, "Implementation of the WECC composite load model for utilities using the component-based modeling approach," in *Proc. IEEE Power Eng. Soc. Transmiss. Distrib. Conf.*, May 2016, pp. 1–5.

[32] D. Karlsson and D. J. Hill, "Modelling and identification of nonlinear dynamic loads in power systems," *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp. 157–166, Feb. 1994.

[33] F. Bai *et al.*, "Measurement-based correlation approach for power system dynamic response estimation," *IET Generat. Transmiss. Distrib.*, vol. 9, no. 12, pp. 1474–1484, Sep. 2015.

[34] B.-K. Choi *et al.*, "Measurement-based dynamic load models: Derivation, comparison, and validation," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1276–1283, Aug. 2006.

[35] Y. Liu, K. Sun, and Y. Liu, "A measurement-based power system model for dynamic response estimation and instability warning," *Electr. Power Syst. Res.*, vol. 124, pp. 1–9, Jul. 2015.

[36] D. Han, J. Ma, R.-M. He, and Z.-Y. Dong, "A real application of measurement-based load modeling in large-scale power grids and its validation," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1756–1764, Nov. 2009.

[37] E. O. Kontis, T. A. Papadopoulos, A. I. Chrysochos, and G. K. Papagiannis, "Measurement-based dynamic load modeling using the vector fitting technique," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 338–351, Jan. 2018.

[38] B.-Y. Ku, R. J. Thomas, C.-Y. Chiou, and C.-J. Lin, "Power system dynamic load modeling using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1868–1874, Nov. 1994.

[39] A. Keyhani, W. Lu, and G. T. Heydt, "Composite neural network load models for power system stability analysis," in *Proc. IEEE PES Power Syst. Conf. Expo.*, vol. 2, Oct. 2004, pp. 1159–1163.

[40] T. Hiyama, M. Tokieda, W. Hubbi, and H. Andou, "Artificial neural network based dynamic load modeling," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1576–1583, Nov. 1997.

[41] M. Bostanci, J. Koplowitz, and C. W. Taylor, "Identification of power system load dynamics using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1468–1473, Nov. 1997.

[42] G. W. Chang, C. I. Chen, and Y. J. Liu, "A neural-network-based method of modeling electric arc furnace load for power engineering study," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 138–146, Feb. 2010.

[43] A. M. Azmy, I. Erlich, and P. Sowa, "Artificial neural network-based dynamic equivalents for distribution systems containing active sources," *IEE Proc.–Generat., Transmiss. Distrib.*, vol. 151, no. 6, pp. 681–688, Nov. 2004.

[44] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[45] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[46] R. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural Comput.*, vol. 2, no. 4, pp. 490–501, 1990.

[47] I. Sutskever, "Training recurrent neural networks," Ph.D. dissertation, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.

[48] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[49] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber, S. C. Kremer, and J. F. Kolen, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Networks*. Piscataway, NJ, USA: IEEE Press, 2001.

[50] K. Greff, R. K. Srivastava, J. Koutnìk, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[51] Z. C. Lipton, J. Berkowitz, and C. Elkan. (2015). "A critical review of recurrent neural networks for sequence learning." [Online]. Available: https://arxiv.org/abs/1506.00019

[52] S. Li, M. Fairbank, C. Johnson, D. C. Wunsch, E. Alonso, and J. L. Proao, "Artificial neural networks for control of a grid-connected rectifier/inverter under disturbance, dynamic and power converter switching conditions," *IEEE Trans. Neural Netw.*, vol. 25, no. 4, pp. 738–750, Apr. 2014.

[53] S. Ahmed-Zaid and M. Taleb, "Structural modeling of small and large induction machines using integral manifolds," *IEEE Trans. Energy Convers.*, vol. 6, no. 3, pp. 529–535, Sep. 1991.

[54] M. Abadi *et al.*, "TensorFlow: Large-scale ma-chine learning on heterogeneous distributed systems," 2016.

[55] A. Pai, *Energy Function Analysis for Power System Stability*. Norwell, MA, USA: Kluwer, 1989.

[56] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

[57] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Feb. 2011.

[58] M. D. Zeiler. (2012). "ADADELTA: An adaptive learning rate method." [Online]. Available: https://arxiv.org/abs/1212.5701

[59] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[60] J. Denecke and I. Erlich, "Dynamic equivalents of active distribution networks," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2017, pp. 1–5.

[61] X.-X. Zhang, Y. Jiang, H.-X. Li, and S.-Y. Li, "SVR learning-based spatiotemporal fuzzy logic controller for nonlinear spatially distributed dynamic systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1635–1647, Oct. 2013.

[62] P. Zhou, D. Guo, H. Wang, and T. Chai, "Data-driven robust M-LS-SVR-based NARX modeling for estimation and control of molten iron quality indices in blast furnace ironmaking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4007–4021, Sep. 2018.

Chao Zheng received the B.S. degree from the School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2014, where he is currently pursuing the Ph.D. degree.

He is currently a joint Ph.D. student with the Department of Energy Technology, Aalborg University, Aalborg, Denmark, supported by the China Scholarship Council. His current research interests include deep learning, computer vision, power system operation, and control.

Shaorong Wang received the B.S. degree from Zhejiang University, Hangzhou, China, in 1984, the M.S. degree from North China Electric Power University, Baoding, China, in 1990, and the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2004.

He is currently a Professor with the School of Electrical and Electronic Engineering, HUST. His current research interests include power system operation and control, smart grid, robotic application, big data, and machine vision.

Yilu Liu (S'88–M'89–SM'99–F'04) received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, and the M.S. and Ph.D. degrees from The Ohio State University, Columbus, OH, USA, in 1986 and 1989, respectively.

She was a Professor with Virginia Tech, Blacksburg, VA, USA, where she led the effort to create the North American Power Grid Frequency Monitoring Network, which is currently operated at the University of Tennessee (UTK), Knoxville, TN, USA, and the Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA, as GridEye. She is currently the Governor's Chair with UTK and ORNL. Her current research interests include power system wide-area monitoring and control, large interconnection-level dynamic simulations, electromagnetic transient analysis, and power transformer modeling and diagnosis.

Dr. Liu is the Deputy Director of the DOE/NSF co-funded Engineering Research Center CURENT and a member of the National Academy of Engineering and the National Academy of Inventors.

Chengxi Liu received the B.Eng. and M.Sc. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 2005 and 2007, respectively, and the Ph.D. degree from the Department of Energy Technology, Aalborg University, Aalborg, Denmark, in 2013.

He was with Energinet.dk, Danish TSO, Fredericia, Denmark, until 2016, and the University of Tennessee, Knoxville, TN, USA, until 2018. He is currently an Associate Professor with the Department of Energy Technology, Aalborg University. His current research interests include power system stability and control, renewable energies, and the applications of artificial intelligence.

Wei Xie received the B.Sc. degree from the Department of Electrical Engineering, Shanghai Jiaotong University, Shanghai, China.

He is currently a Chief Engineer with the State Grid Shanghai Municipal Electric Power Company, Shanghai. His current research interests include smart distribution automation and smart grid.

Chen Fang received the B.Sc. and Ph.D. degrees from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2006 and 2011, respectively.

He is currently an Electrical Engineer with the Electric Power Research Institute, State Grid Shanghai Municipal Electric Power Company, Shanghai, China. His current research interests include renewable energy integration of smart grid and power storage technology.

Shu Liu received the B.Sc. and master's degrees from the Department of Electrical Engineering, North China Electric Power University, Beijing, China, in 2009 and 2012, respectively.

She is currently an Electrical Engineer with the Electric Power Research Institute, State Grid Shanghai Municipal Electric Power Company, Shanghai, China. Her current research interests include microgrid and power storage technology.